Moral Hypocrisy, Power and Social Preferences

Aldo Rustichini, Marie Claire Villeval

GROUPE D'ANALYSE ET DE THÉORIE ÉCONOMIQUE LYON - ST ÉTIENNE

W P 1216

Moral Hypocrisy, Power and Social Preferences

Aldo Rustichini , Marie-Claire Villeval

Revised in August 2014

GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne

93, chemin des Mouilles 69130 Ecully – France
Tel. +33 (0)4 72 86 60 60
Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France
Tel. +33 (0)4 77 42 19 60
Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : gate@gate.cnrs.fr
Téléchargement / Download : http://www.gate.cnrs.fr – Publications / Working Papers

Moral Hypocrisy, Power and Social Preferences

Aldo Rustichini and Marie Claire Villeval

August 2014

Abstract: We study how individuals adjust their judgment of fairness and unfairness when they are in the position of spectators before and after making real decisions, and how this adjustment depends on the actions they take in the game. We find that norms that appear universal instead take into account the players' bargaining power. Also, individuals adjust their judgments after playing the game for real money, when they behaved more selfishly and only in games where choices have no strategic consequence. We interpret this possibly self-deceptive adjustment of judgments to actions as moral hypocrisy. This behavior appears produced by the attempt to strike a compromise between self-image and payoffs, so as to release oneself of one's responsibility for selfish behavior.

Keywords: Moral hypocrisy, fairness, social preferences, power, self-deception, experiment.

JEL Classification: D03, D63, C91, C7

Aldo Rustichini, Department of Economics, University of Minnesota, 1925 4th Street South, 4-101 Hanson Hall, Minneapolis, MN 55455-0462, U.S. E-mail: arust@econ.umn.edu

Marie Claire Villeval, Université de Lyon, F-69007; CNRS-GATE, 93 Chemin des Mouilles, F-69130, Ecully, France. E-mail: villeval@gate.cnrs.fr

Acknowledgments: The authors are grateful to R. Benabou, J. Brandts, G. Charness, H. He, D. Houser, K. Murnighan, D. Nosanzo, A. Poulsen, I. Rodriguez-Lara, D. Zizzo, participants at the Interdisciplinary Workshop in Behavioral and Decision Science at Nanyang University in Singapore, the Symposium on Identity and Incentives in Organizations at Toulouse School of Economics, the conference on Deception, Incentives and Behavior at the University of California at San Diego, the GATE-SEBA workshop in Beijing, the ANR workshop on Conflicts in Rennes, and seminars at the Universities of Amsterdam, Besançon and East Anglia for useful comments. We thank R. Zeiliger for programming the experiment. Financial support from the *Agence Nationale de la Recherche* (ANR BLAN07-3_185547 "EMIR" project and ANR-EMCO11 "HEIDI" project) is gratefully acknowledged. This research was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). AR thanks the NSF, grant SES-1061817.

1. INTRODUCTION

In public, most individuals promote social norms based on fairness and derive utility from being perceived as fair (Benabou and Tirole, 2006; Andreoni and Bernheim, 2009). Parents encourage children to be generous; politicians emphasize dedication to serving others; businessmen promote corporate social responsibility. Real behavior may, however, reveal a different side of human nature. For example, individuals avoid beggars by changing the side of the street; they destroy the resources or production of others by envy (Mui, 1995; Charness *et al*., 2014) or for the joy of destruction (Zizzo and Oswald, 2001); the power of public office leads some politicians to use it for their personal gain (Aidt, 2003). How do people reconcile their stated norms of fairness and the temptation of more selfish actions that may alter their perception as fair people?

In this paper, we study how individuals try to maintain consistency between fairness judgments and real actions, and how they take into account the situation to adjust judgments on norms that appear universal. We test how much they maintain their image of fairness by adjusting their actions to suit their moral judgments, and instead how much they adjust their judgments and manipulate norms to justify their actions. Thus, the study of this interaction belongs to the study of moral hypocrisy (Batson *et al*., 1997, 1999), defined as the motivation to appear moral to oneself and to others while avoiding the cost of acting morally.

Studying moral hypocrisy from an economic perspective is important for several reasons. First, it helps understand how moral reflection influences economic behavior. We may assume that behavior follows given social norms. Here instead we investigate an aspect of the construction of norms by using a dynamic perspective. Studying moral hypocrisy means investigating one aspect of the co-construction of judgments and actions focusing on norm manipulation. Second, studying moral hypocrisy contributes to the understanding of the role of self-image in economic behavior (Bernheim, 1994; Bodner and Prelec, 2003; Benabou and Tirole, 2006; Ellingsen and Johanesson, 2008; Ariely *et al*., 2009). In our experiment we analyze an intrapersonal game by which self-serving individuals may engage, possibly unconsciously, in self-deception to reconcile their judgments with their behavior and keep a good perception of the kind of person they are. Finally, understanding moral hypocrisy contributes to explain why pro-social behavior is not more developed and which incentives could encourage pro-sociality. If greedy individuals who care about looking fair use hypocrisy to keep a high self-esteem while reaping the monetary benefits of acting selfishly,

increasing the cost of moral hypocrisy and norm manipulation may encourage them to behave less selfishly because of reputational concerns

We adopt the methodological view that both actions and judgments constitute social behavior. We expect that individuals strike a compromise between monetary consequences and reputation, understanding that judgments are evaluated in view of actions, and actions are interpreted in view of judgments. Thus, to understand social behavior we need to measure both. Precisely, in our experiment, individuals participated in two consecutive sessions. In the first session we elicit their judgments regarding the fairness and unfairness of all possible shares between two hypothetical players in three different scenarios. The scenarios correspond to Dictator, Ultimatum and Trust games. Eliciting the individual's judgments of fairness and unfairness in terms of intervals rather than point estimates indicates the individual's moral vagueness and constitutes a novelty of our approach. A second novelty is that, instead of eliciting judgments in the position of impartial third parties who would have to make decisions for others (as in Konow, 2000; Konow *et al*., 2009, for example), participants have to report their judgments in the position of partial spectators.[1] They evaluate the fairness and unfairness of all possible shares in the perspective of an advantaged player and a disadvantaged player, successively, without knowing that they will have to make real decisions for themselves in the future. This information allows us to open a different perspective on the study of the effect of strategic environments on stated social norms: eliciting partial judgments helps understanding whether power influences the perception of justice, when we consider the weaker side. Thus realism, as well as selfishness and self-deception, may influence social norms. We find that this is the case.

One week later, the same individuals are invited to play the Dictator, Ultimatum, and Trust games for real. In order to analyze whether actions motivate individuals to adjust their judgments, after making their decision subjects have to again report a judgment on the fairness and unfairness of all possible shares in the same three scenarios. This design allows us to measure both i) how the actions in the second session diverge from the fairness judgments and hypothetical choices during the first session and ii) whether the fairness judgments during the second session conform more to

[1] Our approach differs from philosophical theories of justice based on an ideal observer in the position of a hypothetical third party, or from a judicial spectator theory in which judgments are made by real third parties (Hume). It differs from Rawls's contractarian model of impartiality, in which individuals in original positions choose distributive rules under a veil of ignorance, knowing that they will be stakeholders in the same situation (on the impartial spectator in theories of justice, see Konow, 2009).

the initial statements or to the actions taken. We are also able to determine iii) whether actions are more consistent with initial fairness judgments when the other player is not passive (i.e. in the Ultimatum game) than when he cannot react (*i.e*. in the Dictator and the Trust games).

We observe clear evidence of moral hypocrisy: later actions violate fairness judgments made before individuals know their role, and norms stated later are manipulated in the direction of the actual decisions by more selfish players. The first findings alone would not characterize moral hypocrisy, as good faith in reporting judgments could simply be followed by a lack of willpower. But in most cases, individuals increase the range of shares declared fair after playing the game for real money compared with their initial statement. The discrepancy between hypothetical and real behaviors is larger in games where real behavior has no strategic consequence (Dictator and Trust) than in games where the other player can react to the decision (Ultimatum), so the subject has to take into account consequences of his choices. While the fairness statement in the Dictator scenario is based on an ethical judgment, the strategic dimension of the Ultimatum is immediately perceived (see also Charness and Gneezy, 2008). By using scenarios that vary the bargaining power we show that the adjustment of judgments is influenced by relative power and that both sides, the one in the advantageous and the one in the disadvantageous position, accept the fact that allocations are biased in favor of the powerful.

Moral hypocrisy is not uniform among individuals: those who adjust their judgments to their action the most are also those who behave more selfishly, and whose hypothetical and actual decisions differ the most. This confirms that although it is rational to make selfish decisions from an economic point of view, individuals care about their self-image. By being hypocritical, they both pursue their self-interest and try to keep up appearances of pro-social motivations. However, the most selfish players do not adjust their judgments sufficiently to judge their decision as being fair, having to assume the violation of the fairness norm.

In the remainder of this paper, Section 2 reviews related literature. Section 3 describes the experimental design and procedures. Section 4 presents the hypotheses we test, and Section 5 develops the results. Section 6 discusses our results and concludes.

2. RELATED LITERATURE

The main hypothesis driving our research is that both power and self-serving motives intrude on abstract principles. The idea that power may affect moral judgments is usually presented as the

strong side translating its advantageous bargaining position into a favorable allocation. The idea that power may also affect the opinion of the weak side, making concessions suggested by the advantageous position of the other side seem fair, appears to be new. The notion of moral hypocrisy relates to existing concepts. In social psychology this notion has been conceptualized by Batson *et al*. (1997, 1999) to explain the discrepancies between the norms held by individuals and their actions (see also Stone *et al*., 1997; Watson *et al*., 2006). Individuals want to appear moral without bearing the cost of morality. An underlying mechanism is the individuals' tendency to relieve themselves of their responsibility (Bandura, 1996), which involves self-deception (an "active misrepresentation of reality to the conscious mind" according to von Hippel and Trivers, 2011). Individuals try to convince themselves that serving their own interests does not violate their principles (Trivers, 2011). Our originality is to study moral hypocrisy with standard economic games with real monetary incentives that leave the participants with no possibility to relieve themselves from the responsibility of their actions. Indeed, our design discards strategic ignorance (Dana *et al*., 2007; Grossman, 2010) and strategic avoidance (Lazear *et al*., 2012; DellaVigna *et al*., 2012) as an excuse for selfish behavior in moral dilemmas. It also differs from Haisley and Weber (2011) who show that in a dictator game dictators are more likely to choose the unfair action when the recipient's allocation depends on an ambiguous lottery than on a risky lottery. Lonnqvist *et al*. (2014) address experimentally the difficult issue of whether moral hypocrisy is motivated by management of others' impression or self-deception.

Although related, moral hypocrisy must be distinguished from previous studies of *self-serving biases* (Loewenstein *et al*., 1993). In Babcock *et al*. (1995), players are either under the veil of ignorance or know their actual role when assessing the fairness of bargaining settlements. The authors show that when subjects are not under the veil of ignorance, the two sides consider as fair a position closer to their own direct interest, explaining the frequency of bargaining failures. Similarly, Roth and Murnighan (1982) show that subjects who bargain over how to distribute lottery tickets with asymmetric payments almost doubled the disagreement rate when they knew which side of the bargain they were on (see Murnighan and Pillutla, 1995 for a survey). Our approach is different because we study how partial spectators adjust the judgments of fairness that they report *before* and *after* making decisions as implicated stakeholders. Self-serving biases explain that judgments depend on the information of players on their actual position in the game; moral hypocrisy captures instead the fact that people adjust their fairness judgments to their actions depending on their decisions and on their bargaining power.

Moral hypocrisy can also be related to the notion of *cognitive dissonance* (in psychology, see Festinger, 1957; in economics, Akerlof and Dickens, 1982; Rabin, 1994; Konow, 2000; Oxoby and Smith, 2012). Facing a difference between desires for fairness and self-interest, individuals may reduce this dissonance by adjusting their principles or their actions for the sake of consistency. Mixing theories of equity and of cognitive dissonance, Konow (2000) proposes that the player's objective function includes a material payoff, a cost of dissonance and the cost of self-deception that is used to reduce the dissonance between principles of fairness and self-interested behavior. Indeed, self-deception may allow an individual to choose to believe that it is fair to keep more than the fair share in a dictator game. This model is tested by means of a double dictator game in which dictators make a first allocation decision that affects their earnings and a second decision that determines the share of payoffs between two other players. The observed difference between the two decisions reveals the dictators' judgment on fairness and confirms the prevalence of self-deception. Like Konow (2000), we try to identify cognitive dissonance. In contrast to his approach that identifies self-deception by comparing actions, we compare two statements and two decisions in several games. We analyze the desire to appear moral while avoiding the cost of being moral by studying how judgments on fairness are modified by the action, whereas Konow (2000) does not identify judgments before actions. It should also be noted that moral hypocrisy does not necessarily require inconsistency (Monin and Merritt, 2010).

Measuring moral hypocrisy requires being able to identify people's social preferences. Our analysis requires identifying the fairness judgments separately from the observation of actions. This is in contrast with most analyses that identify social preferences through the observation of incentivized behavior when the latter differs from the equilibrium play determined under selfish preferences. The assumption is that people deviate because of fairness principles (Ostrom, 2000) driven by distributional concerns (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), a concern for efficiency (Charness and Rabin, 2003), or conditional cooperation (Fischbacher and Gächter, 2012). Frohlich *et al*. (2004) and Cappelen *et al*. (2007) have identified a multiplicity of fairness principles. Konow (2000) and Konow *et al.* (2009) have shown that norms depend on a feeling of entitlement. In most of these studies, social norms are, however, identified incompletely, as a single decision is usually observed. Assuming that this decision delimitates the value below which all the other decisions are judged unfair by the individual and above which all decisions are judged fair, would be arbitrary. We differ from this approach by measuring intervals of (un) fairness and by assuming that actions and judgments influence each other dynamically.

Fairness preferences are also sometimes elicited by means of hypothetical questions (Schokkaert and Lagrou 1983; Amiel and Cowell, 1992; Gaertner, 1994; Cappelen *et al*., 2011). In line with the view of impartial observer in philosophical research on justice, individuals are sometimes asked to assess the action of others, not theirs (Falk and Fischbacher, 2006; Cubitt *et al.*, 2011). However, it is unlikely that judgments on the fairness of others' actions define the benchmark used by individuals to evaluate their own action. Indeed, it has been shown that one evaluates more negatively the moral transgression of fairness norms when this transgression is enacted by others rather than by oneself (Valdesolo and deStefano, 2008; Croson and Konow, 2009). Cappelen *et al*. (2011) combine hypothetical and behavioral measures of social preferences in a dictator game. They show that asking subjects to reflect on fairness before making allocation decisions impacts both the assessment of fairness ideals and further distributional choices. Most choices correspond to the self-reported fairness ideal and misreporting cannot be explained by a self-serving bias and it is not correlated with the subject's fairness type. Like them but with a different protocol and more games, we compare judgments and actual behavior. In contrast to them, we elicit judgments again after actual choices have been made because we assume that self-reported preferences can be modified *ex post* by actual choices as a self-deception strategy.

3. THE EXPERIMENT

3.1. Experimental design

The experiment consists of two consecutive sessions, separated by at least one week, both using three different scenarios. We use a within-subject design to observe the judgments and actions of the same individuals placed both in strategic and non-strategic situations.

Scenarios and structure of the sessions

The first scenario corresponds to the Dictator game; the second to the Ultimatum game; the third to the Trust game. In the Dictator game, participant A receives an endowment of 10 points and decides how many points he is willing to transfer to B, while B is passive. A earns the difference between his endowment and his transfer and B earns the points transferred by A. In the Ultimatum game, A receives an endowment of 10 points and he also decides how many points he transfers to B; but in contrast with the previous game, B decides on whether he accepts or he rejects A's offer. If the offer is accepted, A earns the difference between his endowment and his transfer to B, while B earns the points transferred by A. If A's offer is rejected, both participants earn nothing. In the Trust game,

both A and B receive an endowment of 5 points. A decides how many points he transfers to B. The amount transferred is tripled. B decides how many points he sends back to A, between 0 and three times the amount sent by A plus 5 points. A earns the difference between his endowment and his transfer to B plus the amount sent back by B. B earns his endowment augmented by three times the amount transferred by A minus the points sent back to A.

We use these scenarios to elicit the subjects' judgments regarding the fairness and unfairness of all possible transfers in session 1 and at the end of session 2. Precisely, in part 1 (Dictator game) of the first session, the subjects are requested to put themselves in the position of participant A. They evaluate the fairness of all possible shares transferred by A to B, then the unfairness of all possible shares. This reveals their ethical values. After reporting their own judgment, the subjects were asked to indicate which shares they believe most people consider as being fair and unfair. This reveals their empirical and normative expectations: empirical expectations indicate whether the individual believes that a large subset of the population conforms to the norm, and normative expectations indicate whether he believes that a large subset of the population expects him to conform (Bicchieri, 2006). Then, we asked what choice they would make if involved in a real game and which share they believe most people would transfer in the same position. Indeed, when they make a hypothetical choice, subjects may take into account their expectations about what other people would do in the same situation. Next, we asked them to consider the situation from the point of view of the passive player B and to evaluate the fair and unfair shares, and to report their expectations about others' answers. In part 2 (Ultimatum scenario), the structure of the decisions is similar. Subjects report the sets of shares they consider as fair and unfair, in the shoes of the participant A, then in the position of B. They report their expectations about others' answers, make a hypothetical choice, and report their belief about others' choice.

In the first scenario of part 3 (Trust scenario), A transfers 1 point out of 5 to B and keeps 4 points for himself; thus, B can send back between 0 and 8 points to A. In the second scenario, A transfers 4 points out of 5 to B and keeps 1 point for himself; thus, B can send back between 0 and 17 points to A. Participants have to judge the fairness and unfairness of all possible amounts returned by B, first in the position of player B and next in the position of player A. They also make a hypothetical choice in each scenario as player B. Like in the other scenarios, they also report their beliefs about others' norms and actions. In the first session, players are not aware that they will have to play the corresponding games in the second session.

In the second session, participants play the Dictator game (part 1), the Ultimatum game (part 2) and the Trust game (part 3) for money, and again provide fairness judgments in the three corresponding scenarios (part 4).[2] Part 4 replicates the three parts of the first session. This allows us to measure the evolution of judgments depending notably on the actual role and decisions made in each game, while controlling for expectations (we acknowledge, however, that we cannot control for the pure effect of experience). Roles in the game are assigned randomly. The Ultimatum game is played with the strategy method. The B participants are not told the actual choice of their co-participant A before the end of the session, so they have to decide whether to accept or reject each possible offer made by A. Similarly, the Trust game is played under the strategy method. The B participants are not told the amount actually sent by their co-participant A. Thus they have to decide how many points they are willing to return to A for each possible amount sent by him. These games are played one-shot with a random re-matching of participants and roles (A or B) after each game. Using the strategy method allows us to observe the decisions of each individual in reaction to all possible choices of the other player, even those that are made very occasionally. It also prevents decisions from being affected by the preferences of another player and it ensures that they remain psychologically "cold" (see Brandts and Charness, 2010) Since self-reflection on each role has been made important in the first session, we can reasonably assume that using the strategy method did not affect decisions dramatically.

A crucial aspect of our design is that the two sessions were separated by at least one week. We did not remind players of their initial judgments because we did not want the participants to feel committed to follow these judgments due to a consistency bias induced by the design. Choosing a longer time interval between the two sessions would have increased the risk of losing participants and the role of time preferences since payment was made only at the end of the second session.

Elicitation of judgments

To explain what they consider fair and unfair allocations, we let participants communicate intervals rather than point values. These reported intervals constitute what we call "judgments". Specifically, we ask them to answer the following question in the position of player A: *"What do you consider*

[2] The order between the three scenarios has been kept constant across sessions. This does not allow us to control for a possible order effect but the sequence permits a progression in the degree of difficulty of the games, which has facilitated participants' understanding. At the beginning of part 4 in session 2, the three scenarios are reminded all together, so the order of each scenario should not matter. Also, in session 2 judgments are elicited after the three games have been played for money and not after each game. This prevents participants from adjusting their decisions in the next games knowing that they will also have to report their fairness and unfairness judgments.

as being fair shares between A and B?"[3] On the computer screen a bar with two cursors, graduated from 0 to 100%, and a box detailing the choices are displayed (see examples in Fig. 1). Moving the left cursor indicates the minimum fraction going to player B that is considered fair; moving the right cursor indicates the maximum value (indeed, it may be considered as no longer fair to transfer a high percentage of the initial endowment to the passive player).

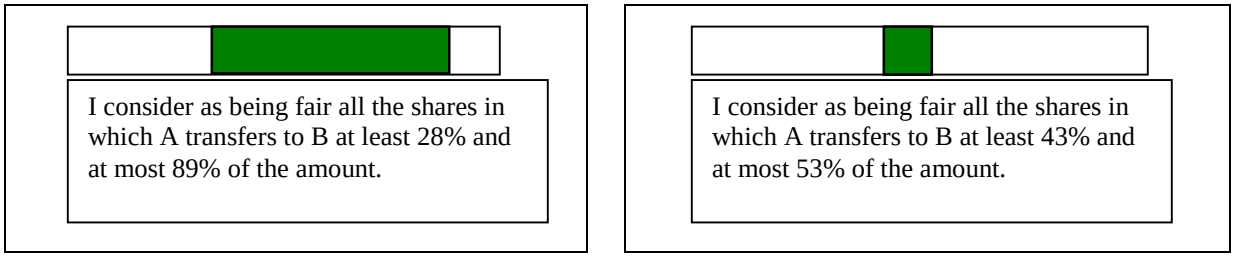


Figure 1. Determination of the sets of fair shares - Two examples of players' screens

Next, the subjects (still in the position of player A) have to evaluate the unfairness of the shares transferred by A to B in the same scenario. A bar, graduated from 0 to 100%, and a box detailing the choices are displayed on the screen (see examples in Fig. 2). By moving the left cursor, the subjects indicate the share below which shares are considered as unfair, and by moving the right cursor they indicate the share above which they are judged as unfair. There may be some overlapping between judgments of fairness and unfairness and some shares may be considered as neither fair, nor unfair; we allow subjects to be inconsistent or indifferent about some shares.

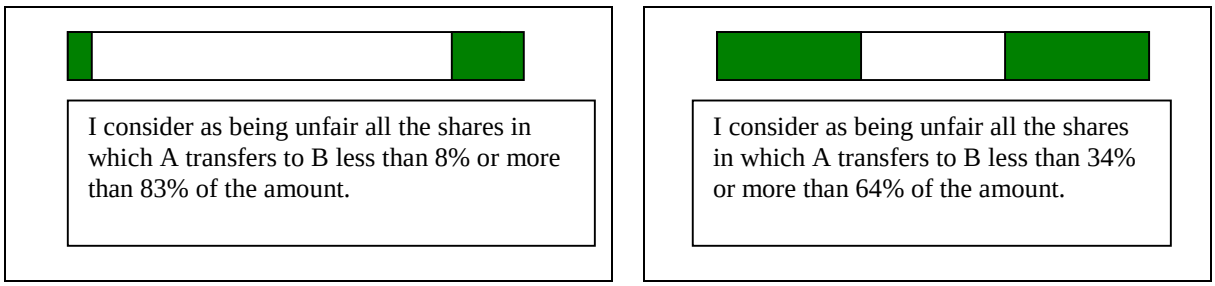


Figure 2. Determination of the sets of unfair shares – Two examples of players' screens

There are several reasons for this design. First, in contrast with what is often assumed by distributive models of inequity aversion, individuals may not view a fair share as a single precise value. More likely they think there is a range of acceptable values. We can identify both the minimum value for which a share is considered as being fair (indicating a limit for advantageous inequity version), but also the value above which shares are no longer fair because they give an

[3] The French original text of the instructions is "*Qu'est-ce que vous considérez comme des partages justes entre A et B ?*". We did not use the alternative word "*équitable*" that evokes equality and is probably less neutral.

excessively high share to the weaker player (giving a limit to disadvantageous inequity version). Second, while this method does not prohibit subjects from reporting a single value if they like; the bounds and the width of the intervals inform on the individual's moral vagueness. It also avoids the problem of focal points associated with point values and it gives the individuals more discretion to make their judgments evolve. Third, it is more difficult for the subjects to remember the bounds of fairness and unfairness intervals from one week to the next. This limits the risk of anchoring effects when we ask them to make decisions and judgments in the second session. Finally, this method should limit the risk of an experimenter demand effect compared to the report of a point value.[4]

This last point is particularly important, as the judgment elicitation was not incentivized. Indeed, the success of our design relies crucially on the assumption that individuals report truthfully their fairness judgments but that these judgments are flexible. Due to lack of incentives, subjects may possibly pay less attention to the quality of their responses. This is also problematic if individuals report a mixture of beliefs and what they believe is expected by the experimenter. As indicated in footnote 4, we did our best not to influence participants and we have the feeling that they reported their preferences very seriously. However, incentivizing judgment elicitation would have been problematic because *i)* it is hard to see what the benchmark would be for the answer to be paid; for example, paying for subjects to be close to the mean response of the subjects in the pool would result in a study of conformism rather than in judgment of fairness; *ii*) it would have motivated people to behave in conformity with their reported judgments while the possible discrepancy is part of what we are studying, and *iii*) it would have introduced differences in payoffs between subjects in session 1, which could have also influenced behavior in session 2.

3.2. Procedures

The experiment consists of 5 sets of 2 sessions each conducted at the laboratory of the *Groupe d'Analyse et de Théorie Economique* in Lyon, France. Undergraduate students from very selective engineering and business schools were invited via the ORSEE software (Greiner, 2004). Individuals

[4] We cannot totally rule out the existence of such effects, motivated by social desirability. However, as for example Zizzo (2010) points out, what is crucial is the belief that the subjects hold about the experimenters' objectives and how they can be linked with the real objectives. Very likely, our subjects were not able to identify the aim of our study, especially in the first session. First, they had to engage in a great number of tasks, making the environment multi-dimensional and our objective not transparent. Second, the time elapsed between the two sessions and the method used for eliciting judgments make it unlikely that the subjects could remember their initial judgments when making decisions, even if they thought that we were studying the connection between the two sessions. We also observe that the demand effect here works against the hypothesis we suggest: if subjects thought that the experiment was about fairness, strong demand effects should have driven behavior in the direction of initial judgments, not the opposite.

committed to participate in two sessions and they were informed that the earnings made in the two sessions would be paid only at the end of the second session. 83 individuals (52% were females) participated in the two sessions.[5] None of them had participated in a bargaining game or in a Trust game before. The experiment was computerized using the REGATE platform.

Upon arrival at the first session, the subjects extracted a tag from a bag indicating their computer user name and received another tag with a password. They were told that they had to bring back this tag for the second session to be allowed to participate.[6] They were informed that they would be paid €8 at the end of the second session for participating in the first session. They received sets of instructions for each scenario (see Appendix A) after completion of the judgments and hypothetical decisions in the previous scenario. They answered a comprehension questionnaire; understanding was checked individually and questions were answered in private. They were informed at the beginning of the first session that their responses would in no case influence their participation in the second session or the content of the second session. At the beginning of the second session, the subjects entered their password in the computer. In each part, after a new check of their understanding, we paired them and assigned randomly the two roles. It was made public in the instructions (see Appendix B) that the decisions would not be communicated to the other subject until the end of the session. Then, we distributed a final set of instructions reminding the participants about the three scenarios. Finally, the subjects received feedback on the decisions of their co-participants in the first three parts of the second session.

Participants were paid €8 for participating in the second session in addition to the payment for one of the first three parts of the session that was randomly drawn at the end of the session. A secretary who was not aware of the content of the experiment made payment in a separate room. This fact was made public to all subjects from the very beginning of the first session. The first session of the experiment lasted on average 75 minutes and the second, 90 minutes. Each participant earned an average of €28.99 (standard deviation: €8.18).

[5] Five subjects did not show up at the second session. One subject showed up in the second session and not in the first one; he was nevertheless accepted since we needed an even number of participants. The data from these six subjects are not included in the data analysis.

[6] The composition of the groups differ between the two sessions, as participants registered at the same time for two sessions but chose among various schedules of each session. We did this on purpose to limit communication between subjects. This is also why we dismissed people at different moments at the end of the first session.

4. SOCIAL NORMS AND HYPOCRISY: PREDICTIONS

What is a natural benchmark for abstract norms? Consider the Dictator game. Fairness may be defined as equal splitting, as long as there are no reasons to think that the position of Dictator deserves special treatment because it is earned or deserved. The Rawlsian principle of justice as fairness assumes that one should not take advantage of a position of greater power, or accept the consequences of a position of weaker power. Equality of payoffs is the reference standard in distributional models (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Falk and Fischbacher, 2006). However, when this ideal translates into actual proposals, the advantageous bargaining position might creep into the moral reasoning and bias the actual proposal in favor of the Dictator. In the Ultimatum and Trust games, the same benchmark should apply when the strategy space is unlimited, such as in our experiment. When confronted with actual choices, other motives may impact the judgment of fairness, such as efficiency and reciprocity (see the intention-based models of preferences such as Rabin, 1993, Dufwenberg and Kirchsteiger, 2004). We take as the null hypothesis the Rawlsian idea that individuals use the equal splitting rule to make their judgment of fairness and unfairness, independently of their role.

Hypothesis 1: *When the strategy space is unrestricted, equal split of payoffs is the reference point defining fairness and unfairness judgments in the perspective of both players. We test this hypothesis in section 5. 1.*

When real behavior is not readily available for comparison, judgments and hypothetical behavior are not constrained to some measure of consistency with actual choices, and are free to follow ideal positions that may reflect well on those who provide them. Instead, when the check provided by real behavior is possible, the cost attached to the failure to implement a self-flattering judgment is the bad signal of being inconsistent. We predict that when real behavior can be used to judge the consistency of facts and statements, they are more prudently in tune with actions.

Hypothesis 2: *Stated fairness judgments will be farther from equal allocation after subjects have made actual transfer choices that violate those judgments, moving the stated norm closer to actual behavior. We test this hypothesis in section 5. 2.*

Our next focus is on the relationship between the expressed social preferences, strategic reasoning and behavior. We hypothesize that a strategic thinking mode may or may not be activated as players contemplate the game and are cued to view the situation only as an abstract moral setting

or also as a strategic setting. The three games differ in important respects in the way ethical and strategic considerations interact at the moment of formulating moral judgments. In the Ultimatum game, the moral evaluation of the sender's proposal may have practical consequences because he knows that the receiver can reject the offer and decrease efficiency. Thus, the Ultimatum game naturally cues a strategic view even when the norm is stated as abstract. This was clear in our setting since this scenario was presented to the players just after the Dictator scenario, making the second mover – and therefore the strategic dimension of the game - more salient. The other two scenarios do not cue strategic thinking in the abstract setting since evaluations are void of practical consequences. In both the Dictator and Trust games, the evaluation given by the receiver has no consequence because he cannot act to affect the outcome in any way.

The strategic cue is instead active when participants judge the fairness of moves ex-post. In our second session, when the individual is asked to choose and then to provide a fairness judgment, he considers the question of the fair allocation together with the practical implication of the allocation among people that he just witnessed. The practical consequence now has more weight and the fairness judgments reflect the power and the intention of the first mover. In the Ultimatum scenario this is not new because the real implication of the judgment has been present from the start, so we hypothesize that in this case the change in the judgment will be smaller.

Hypothesis 3: *Real choices are more selfish than hypothetical ones because the environment of decision-making is different. The discrepancy between the two will be larger in games where real behavior has no strategic consequences (Dictator and Trust games) than in games where consequences are possible (Ultimatum game). We test this hypothesis in section 5. 3.*

Finally, we test the hypothesis of moral hypocrisy, *i.e*. the fact that the individuals who behave differently in hypothetical and real situations adjust their judgments to better match their actions. This should hold regardless of the subjects' expectations about others. The adjustment is an essential part of the hypothesis. Taken separately, the facts that hypothetical and actual choices differ and that behavior departs from fairness judgments are not sufficient to identify hypocrisy. Indeed, one may behave selfishly and consider that it is the "right" thing to do, while reporting that fairness requires equal sharing. Indeed, rightness is a notion that encompasses fairness and other principles (Konow, 2001). Moreover, according to Bicchieri (2006), individuals make choices based on their expectations about what others do and expect them to do, while fairness judgments are based on a different set of information (personal ethical values, fairness ideals). Finally, actions may differ from

initial fairness judgments because good faith in reporting fairness judgments in the first session could be followed by a lack of willpower.

This adjustment will likely differ across individuals, depending on the generosity of their actual choice and on their bargaining power in the games. Individuals are expected to adjust their judgments more when they make more selfish actions in scenarios where they have the last word (Dictator and Trust scenarios).

Hypothesis 4: *The likelihood of adjustment of fairness and unfairness judgments to actions is more likely to be observed when individuals make more selfish real decisions in scenarios where they have a stronger bargaining power. We test this hypothesis in section 5. 4.*

5. RESULTS

5. 1. The fairness judgments

Our data show that fairness or unfairness judgments take into account the relative position of the two sides. This is true independently of the point of view participants are asked to take. Table 1 displays the mean lower and upper bounds of the fairness and unfairness judgments in each scenario and from the perspective of each player, as reported in the first session. Mean bounds are expressed in percentage points of the endowment transferred to the receiver in the Dictator and Ultimatum scenarios, and in percentage points of income (equal to 3 times the amount received from the sender plus 5 points of initial endowment) returned to the sender in the Trust scenarios.

(*Insert Table 1 here*)

Table 1 shows that in the first session the mean upper and lower bounds of stated fair shares are clearly in an asymmetric position with respect to an equal share. In the Dictator scenario, the midpoint of the upper and lower values for the dictator's point of view on fairness is 40.92% of the initial endowment, which is significantly different from 50% (*t*-test, two-tailed, $p<0.001$).[7] In the Ultimatum scenario, the midpoint of the upper and lower values for the sender's point of view is 41.58%, which is also significantly different from 50% (*t*-test, $p<0.001$). In the Trust scenarios, an equal share between the two players would require a return of 25% of total income in case of a low transfer by player A (corresponding to a return of 2 points) and of 47% of total income in case of a

[7] All *t*-tests and Wilcoxon signed-rank tests used in this paper are two-tailed *t*-tests with the individual as the unit of observation, unless specified otherwise. We use midpoints as one of the metrics of fairness judgments but one must keep in mind that midpoints may not correspond to the "most fair allocation" in the view of subjects because of asymmetric inequity aversion.

high transfer (a return of 8 points). The midpoint of the upper and lower values for the receiver's point of view is 43.16% in case of the high transfer, which is significantly different from 47% (*t*-test, p=0.002). In contrast, the midpoint is 29.89% in case of a low transfer, which is significantly *higher* than the return rate of 25% required by an equal share (t-test, p<0.001). Several interpretations are possible for this result. In the perspective of the decision-maker this is for example consistent with the notion that utility is more strongly affected by disadvantageous inequality aversion than by advantageous inequality aversion. Of course this interpretation requires making these preferences a criterion for justice. Several of the findings indicated below are consistent with this interpretation.

The same asymmetry holds for the unfairness judgments. The area defined by the lower bound of unfair shares ('*transferring less than'*) is always significantly smaller than the area defined by the upper bound ('*transferring more than'*). Individuals identify more unfair shares at the detriment of the decider than unfair shares[8] at the detriment of the beneficiary of the transfer.[9]

Interestingly, asymmetries are also observed in the perspective of the weaker player, favoring the stronger player. For example, in the Dictator scenario, the lower bounds of fairness judgments do not differ significantly when placed in the shoes of the sender or in that of the receiver (Wilcoxon test – W, hereafter-, p=0.399). The midpoint of the mean upper and mean lower values of fairness is slightly modified but it is also significantly different from 50% in both the Dictator (*t*-test, p<0.001) and the Ultimatum scenarios (*t*-test, p=0.018) in the perspective of the receiver, like it was in the perspective of the sender. Similarly in the Trust scenarios, the midpoint differs from 25% in case of the low transfer (*t*-test, p<0.001) and from 47% in case of the high transfer (*t*-test, p=0.013) from the point of view of the sender. The same asymmetry is observed regarding the unfairness

[8] In the Dictator scenario, players asked to take the point of view of the dictator consider as unfair to transfer less than 18.93% of the endowment on average (S.E.=1.68) and more than 68.67% (S.E.=1.76). In the perspective of the sender in the Ultimatum scenario, the mean lower bound is 20.24% of the endowment (S.E.=1.58) and the mean upper bound is 66.14% (S.E.=1.86). In the perspective of the receiver in case of a low transfer by A in the Trust scenario, the lower bound is 13.40% of total income (S.E.=1.15) and the upper bound is 52.26% (S.E.=2.40); the corresponding values are 25.80% (S.E.=1.72) and 63.86% (S.E.=1.99) in case of a high transfer. Wilcoxon tests comparing the width of the area delimited by the lower bound and the minimum possible transfer and the width of the area delimited by the upper bound and the maximum possible transfer show that all these differences are highly significant (p<0.001).

[9] Since players determine the unfair shares without being reminded which shares they defined as being fair, there may some overlapping between fair and unfair shares. This can be thought of a margin of error. In session 1 (session 2, respectively) we observe overlapping between the lower bounds of fairness and unfairness in 14.46% of the players in the Dictator game (15.66%), 7.23% in the Ultimatum game (2.41%), and in the Trust game 8.43% when the transfer is low (12.05%) and 10.84% when it is high (9.64%).

bands.[10] Players in the weaker position make judgments that anticipate that the stronger side will use its advantageous position due to the endowment effect and his bargaining power, rather than claiming that allocations should be independent of the bargaining position.

Finally, the mean judgments on fairness in the sender's perspective in Dictator and the Ultimatum scenarios are relatively similar. In the perspective of the sender, the mean fairness band is defined by the intervals [24.19, 57.65] in the first scenario and [24.71, 58.45] in the second scenario; the fairness band is slightly shifted to the right but the width of the two bands is not significantly different (W, p=0.929), in spite of the fact that the receiver in the Ultimatum scenario has a veto power while it is passive in the Dictator scenario. This may result from counteracting forces; in the Ultimatum scenario the choice is less perceived as ethical and more as strategic, and so the sender takes into account the possibility that the receiver might reject a too low offer. In the Dictator scenario this strategic element is absent, the judgment is perceived as an ethical statement, and the sender is stating as fair a more generous offer. Not surprisingly, in the perspective of the receiver, the lower and upper bounds of fairness are moved to the right in the Ultimatum compare to the Dictator scenario (not significantly so, however: W, p=0.186 and p=0.140, respectively).

These results do not support our null hypothesis 1:

Result 1: *Abstract judgments of fairness and unfairness concede more to the player with the higher bargaining power. This is true for judgments made from the perspective of both the most and the least powerful players, in both cases favoring the player who has the more to lose.*

5. 2. Behavior and judgments: From real acts to words

Table 2 displays the evolution of mean fairness (panel A) and unfairness (panel B) judgments between the first and the second sessions for each scenario, depending on the player's role in the games played in session 2. Like in Table 1, in the Dictator and Ultimatum scenarios mean bounds are expressed in percentage points of the endowment transferred by the sender. In the Trust scenarios, they are expressed in percentage points of income returned by the receiver.

(*Insert Table 2 here*)

[10] When players take the point of view of the receiver, the width of the lower unfair band is significantly smaller than that of the upper band in the Dictator (W, p=0.003) and the Ultimatum scenarios (W, p=0.090). When they are asked to take the point of view of the sender in the Trust scenarios, the difference is significant in the case of a low transfer but not in the case of a high transfer (W, p<0.001 and p=0.386, respectively).

First, Table 2 shows how more powerful players modify their judgments in the second session when taking the perspective of the sender. In the Dictator scenario, actual senders adjust downward both lower and upper bounds of their fairness statement (see panel A). In the first session on average they consider that 22% of the endowment is the minimum to consider a transfer as being fair and that a share is fair up to 59.44% of the endowment; in the second session, 15.8% is the minimum and 56.34% is the maximum. These differences are significant (W, p=0.012 for the lower bound; p=0.123 for the upper bound). If one considers midpoints of fairness bands, the asymmetry of fair shares relative to the equal split is increased in the second session compared to the first one. The difference between the midpoints in the first and the second sessions is significant in the Dictator scenario (W, p=0.021) and in the Trust scenario when the transfer is high (W, p=0.009). The same conclusion applies to the evolution of the unfairness judgments that are also adjusted downwards (see panel B).

In contrast, the asymmetry of fair shares is not significantly modified in the second session in the Ultimatum scenario where the receiver can still reject the offer (W, p=0.325). The difference is not significant either in the Trust scenario when the initial transfer is low (W, p=0.653). The fact that average judgments are not modified in all scenarios is interesting. We cannot totally exclude that thinking about a decision problem in a given role, in a game played for money, may be more helpful in terms of learning in some scenarios than in others. However, we believe that it is more likely that, from the very beginning, the Dictator and the Trust scenarios are not perceived like the Ultimatum scenario, as suggested by Charness and Gneezy (2008). Moreover, strategic considerations (and a possible concern for efficiency) in this environment reduce the variance of decisions. This does not mean that morals do not play a role in the Ultimatum game, but strategic and moral considerations are certainly weighted differently in these various environments.

Another striking result is that players who have been assigned a weaker role also adjust downwards their fairness judgments when placed in the perspective of the other player compared with their first judgment, except in the Ultimatum scenario. For example, Table 2 shows that in the first session, they consider on average that 26.33% of the endowment is the minimum to consider a transfer as being fair in the Dictator scenario; in the second session, instead they consider on average that 20.74% is the minimum (W, p=0.020). Since the upper bound of fair shares is not modified (W, p=0.550), not only the definition of fairness becomes more asymmetric relative to an equal sharing (the midpoint is lower in session 2 than in session 1, one-tailed W, p=0.079), but also the range of

fair shares is increased after playing the game for money (W, p=0.092). The adjustment of the receivers is, however, smaller than the dictators; thus, the lower bounds of fairness now differ between them (Mann-Whitney U-test, p=0.077). In the Trust game with a high transfer, the senders also adjust downward the lower bound of the fair returns in the perspective of the receiver (W, p=0.059), moving the midpoint to the left (W, p=0.097). In contrast, when placed in the perspective of the sender in the Ultimatum scenario the receivers do not adjust significantly the lower bound of their fairness judgment (W, p=0.542), the lower bound (W, p=0.393) or the upper bound of unfairness judgments (W, p=0.745); they increase the upper bound of fairness (W, p=0.051). In all scenarios except the Ultimatum, the adjustment of judgments by the actual weaker players in the perspective of the stronger players may be due to the fact that the difference in bargaining power becomes more salient since real choices are now being made.

This analysis leads to the following results that support Hypothesis 2.

Result 2a: *In all scenarios except the Ultimatum, on average, actual decision-makers adjust their judgments by relaxing their norm of fairness by extending the range of fair allocations to their advantage.*

Result 2b: *In most scenarios and on average, weaker players also adjust downwards their fairness and unfairness judgments when placed in the perspective of the decision-maker.*

5. 3 Hypothetical and real choices

Table 3 displays the distributions of hypothetical and real choices in the Dictator game (left panel) and the Ultimatum game (right panel). Table 4 reports the distributions of hypothetical and actual amounts returned by the receivers in the cases of both a low (left panel) and a high (right panel) transfer from the sender. Our findings are in agreement with Hypothesis 3: In the Dictator and Trust games, where the evaluation of the action of the other has no consequence, hypothetical transfers in the first session and real choices in the second session differ. In the Ultimatum game, where the evaluation has consequences, the two choices do not differ.

(*Insert Tables 3 and 4 here*)

Consider the Dictator game, unsurprisingly, senders are much less generous when making real choices than hypothetical ones. 34.15% of them play the equilibrium (0) in the second session while only 14.63% of them did so when making a hypothetical choice. The mean actual transfer represents 17.80% (S.E.=2.87) of the endowment against 26.59% (S.E.=2.36) for the hypothetical transfer. The

difference is highly significant (W, p=0.001). In the Trust game, real amounts are also lower than the hypothetical returns and by a sizeable amount. In the case of a low initial transfer, the zero return represents only 7.14% of the hypothetical choices of the receivers but 38.10% of their actual decisions. In case of a high transfer, the corresponding values are 0% and 23.81%. The picture is the same when one considers the percentages of returns that would equalize payoffs between senders and receivers. In the case of a low transfer, 52.38% of the hypothetical choices but only 19.05% of the real choices are equal to 2; in the case of a high transfer, 26.19% of the hypothetical choices but none of the real choices are equal to 8. The average actual return in the second session if the sender made a low transfer is 0.88 (S.E.=0.137) whereas the average hypothetical return was 1.74 (S.E.=0.150); the corresponding values are 4 (S.E.=0.447) and 6.38 (S.E.=0.334) if the sender has sent 4 points. All these values differ significantly (W, p< 0.001).

Instead, hypothetical and real choices are very close in the Ultimatum game. The equilibrium play (0 or 1) represents 7.14% of the hypothetical choices of the senders and 4.76% of their actual decisions. The mean actual transfer to the receivers amounts to 35% of the senders' endowment (S.E.=1.71) whereas the mean hypothetical transfer amounted to 34.29% of their endowment (S.E.=1.84). The difference is not significant (W, p=0.683), which is consistent with the hypothesis that strategic considerations influence both hypothetical and real choices in this game.

To explore the determinants of choices, we report in Appendix C the results of regressions in which the dependent variable is either the hypothetical choice made in each game (Table A1) or the actual choice (Table A2). In Table A1, the independent variables include all subjects' expectations about the hypothetical choices of others, their reported lower and upper bounds of fairness, and their normative expectations about the fairness bounds reported by others. In Table A2, they include the active players' hypothetical choice and their expectations about others' actions in the second session. We find that *hypothetical* choices are significantly influenced by the lower bound of fairness reported by the subjects (except in the Trust game with a high transfer): the higher is the lower bound, the larger is the hypothetical transfer. Hypothetical choices in each game are significantly correlated with expectations about whether others conform to the norm; in contrast, they are not correlated with normative expectations, except in the Dictator game. We also find that the subjects' *actual* choices are always significantly correlated with their hypothetical choices (at the 1% level in all games except in the Ultimatum scenario in which the level of significance is 5%) and, except in the Dictator game, with their expectations about others' actual behavior in the second session

(significant at the 1% level). This suggests indirectly that the subjects' initial reports on fairness intervals were meaningful judgments, although not incentivized.

These analyses support our next result.

Result 3: *In the Dictator and Trust games hypothetical and real choices differ on average, and the change favors the first mover. The difference is smaller in the Ultimatum game. The difference partly depends on empirical expectations about others' behavior.*

5. 4 Actual fairness and adjustment of judgments

In this last sub-section, we complement the previous analysis by studying, at the individual level, whether the players adjust their fairness judgments to their actual decisions and in which circumstances. To do so, we have estimated for each scenario a model in which the dependent variable is the difference in the lower bound of fairness judgments between the second and the first sessions. The independent variables include the subjects' empirical expectations about others' actual choices in the second session, a dummy variable indicating whether the subject has an active role (sender in the Dictator or Ultimatum games, or receiver in the Trust games) or not, and a variable capturing the difference between the actual and the hypothetical decisions when the individual has an active role. People who adopt a hypocritical behavior are expected to adjust their fairness judgments to the discrepancy between their actual and hypothetical choices, beyond any adjustment that would be related to empirical expectations about others' choices. The models are estimated with robust standard errors. Estimates are displayed in Table 5.

(Insert Table 5 here)

As predicted by our hypothesis, Table 5 shows that in all games the size of the adjustment of fairness principles is significantly influenced by the discrepancy between actual and hypothetical choices (at the 1% level in Dictator and Ultimatum scenarios and at the 5% level in the Trust scenarios). The more selfish the actual choice is compared to the hypothetical choice, the more the individual revises his fairness judgment downward compared to his initial judgment. Playing an active role has no additional influence on the revision of judgments. This provides evidence of moral hypocrisy through norm manipulation at the player's advantage. This result is obtained after controlling for empirical expectations about others' choices. Expectations impact the adjustment only in the Dictator scenario; in this scenario, expecting higher transfers by the others raises the lower bound of fairness judgment between the two sessions (or limits its downward revision).

Interestingly, the coefficient associated with the difference between the actual and the hypothetical choice is 0.418 in the Ultimatum scenario and 0.239 in the Dictator scenario, and between .240 and .274 in the Trust scenarios. Yet, our previous analysis has shown that on average the adjustment of the fairness judgments is much smaller in the Ultimatum scenario. To reconcile these facts, we proceed to an analysis taking also into account the relative values of the actual transfers made by the sender in the Dictator and the Ultimatum games and by the receiver in the Trust game. In each game we divide the sample into three categories that correspond to transfers below the median, equal to the median, and above the median. Table 6 indicates, for each game and each transfer category, the percentage of variation between the mean actual and hypothetical transfers (column 1), the percentage of variation between the mean lower bound of fairness judgments in session 2 and in session 1 (column 2), the percentages of variation between the mean actual transfer and the lower bound of fairness stated in session 1 (column 3) and in session 2 (column 4). Each column includes the *p*-value of Wilcoxon signed-rank tests.

(*Insert Table 6 here*)

Table 6 delivers three main findings. First, the evolution of fairness judgments is never significant for the players whose transfers are above the median, even when these decisions are more generous than the hypothetical choices (as in the Ultimatum game). In this situation, there is no need to manipulate the judgment since the decision lies within the fairness interval. Second, Table 6 identifies situations in which the percentage of evolution of the fairness judgment is similar to the difference between the actual and the hypothetical choices. In these situations, the subjects play more selfishly than in the hypothetical environment and they manipulate their judgment to be or to remain above the lower bound of fairness. The third finding concerns the individuals who play the most selfishly by making transfers below the median in the Dictator and the Trust game. These subjects are characterized by a very large discrepancy between their actual and hypothetical choices. They adjust their judgments downward considerably, but not sufficient enough, however, to allow their decisions to be included in their revised fairness interval.[11] Many of them evade their fairness judgment, as the manipulation of these judgments needs to be extremely large to allow them to conform to it. This situation is only observed in the scenarios where individuals have the higher bargaining power.

[11] In the Trust game with a low transfer, judgments are not even modified significantly. In contrast to the other games, a low return can be justified by a reciprocal reaction to the first mover's low level of trust. We also acknowledge that in this context there are less possible returns and therefore lesser possibilities to adjust the fairness bands.

This analysis leads to our last results that confirm Hypothesis 4:

Result 4a: *Controlling for expectations, a higher discrepancy between actual and hypothetical choices leads to a larger manipulation of fairness judgments. We call this discrepancy moral hypocrisy.*

Result 4b: *More selfish behavior of players with a higher bargaining power is associated with a larger manipulation of fairness judgments, but also with more norm evasion.*

6. DISCUSSION AND CONCLUSION

Our results suggest the general rule that the practical implications of actions affect the fairness judgments, and *vice versa*. This intrusion of strategic evaluation into normative settings occurs in two ways.

First, norms that appear abstract take into account the bargaining power of the two sides. Players in the advantageous position anticipate how future behavior might deviate from tight moral standards, and thus choose to make them less stringent beforehand; for example, their judgments deviate from the norm of equal sharing. Power works the other way too: individuals in weaker positions also anticipate that the stronger side will take advantage of the position, and they adjust the moral judgment to this prediction, declaring as fair unequal allocations that favor the opposite side. Thus, power intrudes in ethical judgments, in the mind of the weak and the strong, and bends the norm in favor of the powerful.

Second, people employ principles of justice in self-serving ways (Konow, 2000; Cappelen *et al*., 2007, 2011, Babcock *et al.*, 1995; Rodriguez-Lara and Moreno-Garrido, 2011). Controlling for their beliefs about others' behavior, individuals adjust the range of fair shares and unfair shares after playing the game for real money compared with the initial statement they gave when the criteria of fairness and unfairness were elicited as universal but inconsequential norms. Moral hypocrisy is used as a tool to manage the tradeoff between the immediate convenience of the actions and the conflict these actions create with judgments. It balances the need to maintain a positive self-image and the convenience of a present choice.

There is experimental evidence that people not only judge intentions, independently of outcomes, but that they anticipate others doing so (Rabin, 1993; Charness and Rabin, 2002). But then, if people understand the importance of the evaluation of intentions given by others, it is all the

more important for them to adjust their stated judgments to their actions to send a signal that reassures the others of the fairness of their intentions. The size of the adjustment is reduced when another player has behaved selfishly (i.e. in the Trust game when the sender expressed low trust). The adjustment of judgments to actions is considerably smaller when there are strategic reasons that dictate prudence and fairness in deciding transfers. The discrepancy between hypothetical and real behavior is larger when the action being judged has no further consequence (as the first move in the Dictator and the second move in the Trust game) than when it does (as the first move in the Ultimatum game). If an allocation has the strategic value of affecting future actions, then it is perceived differently already in the first session. This difference in perception is reflected in the lack of adjustment of judgments and choices in the second session, since the action was already evaluated as strategic in the first session, the fact that it has real consequences of affecting payoffs of individuals is not a novelty. The fact that individuals adjust their statements when they are the last movers but not when strategic motives are present shows that the adjustment of judgments in the direction of selfishness cannot be simply interpreted as a learning effect. Similarly, learning cannot explain the diverging changes between hypothetical and actual choices according to the level of transfers relative to the median. We show that more power facilitates the enunciation of generous ethical judgments that may be distant from real actions because individuals do not examine the situation as a strategic one when they express them.

The discrepancy between judgments and acts and the subsequent adjustment to actions of fairness and unfairness statements pose the question of a systematic study of moral hypocrisy in strategic behavior. Beyond the intrusion of expectations about what others think and do in the moral reasoning, occurrence of hypocrisy is due to the fact that people build an identity when stating their initial principles and hypothetical choices. Since there is no cost for looking pro-social when roles are not assigned yet, most of them build an excessively generous image of themselves. However, our design forces people to make decisions in the second session, when acting in accord with pro-social judgments is costly; at that moment they cannot remain ignorant of their true identity as there is no opt-out option. After making choices that are usually less pro-social than the initial judgments, especially if they get the highest bargaining power, they deceive themselves. They adjust their judgments in the direction of a better alignment with their actions to keep up the appearance of being pro-social and to maintain a positive self-image. By reducing the distance between their fairness judgments and their actions, people may convince themselves that their actions do not hurt their morale since they are closer or belong to a more permissive fairness interval. Moral hypocrisy is not

systematic, however, hypocritical players are those who reported initially the most generous statements and who then behaved more selfishly. In contrast, the individuals who behave more generously do not feel the need to reevaluate their judgments.

Many extensions are required to better understand the role of power and moral hypocrisy in ethical judgment. One of these extensions would put participants in the position of impartial spectators making decisions for others, as third parties, after having made their partial judgments from the perspectives of each party. In this environment, eliciting what individuals think one should do would help differentiate judgments on what is right and judgments on what is fair. Finally, identifying separately the conscious and unconscious mechanisms behind moral hypocrisy and the maintenance of self-image would help to better understand its deep motivation and how incentives should be redesigned to encourage pro-social behavior. An important direction of inquiry should clarify how conscious the misrepresentation is. Our design does not allow us to test whether moral hypocrisy is conscious or unconscious. Hypocrisy may be a conscious attempt to claim morality while acting selfishly but it may also be unconscious, based on self-deception (von Hippel and Trivers, 2011).

REFERENCES

Aidt, T.S., 2003. Economic Analysis of Corruption: A Survey. *The Economic Journal* 113, F632-652.

Akerlof, G.A., Dickens, W.T., 1982. The Economic Consequences of Cognitive Dissonance. *American Economic Review* 72(3), 307-319.

Amiel, Y., Cowell, F.A., 1992. Measurement of income inequality: experimental test by questionnaire. *Journal of Public Economics* 47, 3–26.

Andreoni, J., Bernheim, B.D., 2009. Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica* 77(5), 1607-1636.

Ariely, D., Bracha, A., Meier, S., 2009. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review* 99(1), 544-555.

Babcock, L., Loewenstein, G., Issacharoff, S., Camerer, C., 1995. Biased Judgments of Fairness in Bargaining. *American Economic Review* 85(5), 1337-1343.

Bandura, A., 1996. Mechanisms of Moral Disengagement in the Exercise of Moral Agency. *Journal of Personality and Social Psychology* 71(2), 364–367.

Batson, C.D., Kobrynowicz, D., Dinnerstein, J.L. Kampf, H.C. Wilson, A.D. 1997. In a Very Different Voice: Unmasking Moral Hypocrisy. *Journal of Personality and Social Psychology* 72(6), 1335–1348.

Batson, C. D., Thompson, E.R., Seuferling, G., Whitney, H., Strongman, J., 1999. Moral Hypocrisy: Appearing Moral to Oneself Without being so. *Journal of Personality and Social Psychology* 77(3), 525–537.

Benabou, R. Tirole, J.. 2006. Incentives and Prosocial Behavior. *American Economic Review* 96(5), 1652-1678.

Bernheim, B.D.. 1994. A Theory of Conformity. *Journal of Political Economy* 102(5), 841–877.

Bicchieri, C., 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Bodner, R., Prelec, D., 2003. Self-signaling and Diagnostic Utility in Everyday Decision Making. In *The Psychology of Economic Decisions,* eds. I. Brocas and J.D. Carrillo, 105-123. Oxford: Oxford University Press.

Bolton, G.E., Ockenfels, A.. 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 90 (1), 166–193.

Brandts, J., Charness, G., 2011. The Strategy Versus the Direct-response Method: a First Survey of Experimental Comparisons. *Experimental Economics* 14, 375-398.

Cappelen, A.W., Hole, A.D., Sørensen, E.Ø, Tungodden, B.. 2007. The pluralism of fairness ideals: an experimental approach. *American Economic Review* 97(3), 818-827.

Cappelen, A.W Hole, A.D., Sørensen, E.Ø, Tungodden, B., 2011. The importance of moral reflection and self-reported data in a dictator game with production. *Social Choice and Welfare* 36(1), 105-120.

Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117(3), 817-869.

Charness, G., Gneezy, U., 2008. What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, 68(1), 29-35.

Charness, G., Masclet, D., Villeval, M.C.. 2014. The dark side of competition for status. *Management Science* 60(1), 38–55.

Croson, R., Konow, J., 2009. Social preferences and moral biases. *Journal of Economic Behavior and Organization* 69(3), 201-212.

Cubitt, R.P., Drouvelis, M.,Gächter, S., Kabalin, R., 2011. Moral judgments in social dilemmas:

How bad is free riding? *Journal of Public Economics* 95(3-4), 253-264.
Dana, J., Weber, R.A., Kuang, J.X., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory* 33(1), 67-80.
DellaVigna S., List, J.A., Malmendier, U., 2012. Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics* 127(1), 1-56.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47, 268–298.
Ellingsen T., Johannesson, M., 2008. Pride and Prejudice: The Human Side of Incentive Theory. *American Economic Review* 98(3), 990-1008.
Falk, A., Fischbacher, U.. 2006. A theory of reciprocity. *Games and Economic Behavior* 54(2), 293-315.
Fehr E., Schmidt, K.M., 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114(3), 817-868.
Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free-riding in public good experiments. *American Economic Review* 100 (1), 541-556.
Frohlich, N., Oppenheimer, J.A., Kurki, A., 2004. Modeling other-regarding preferences and an experimental test. *Public Choice* 119, 91-117.
Gaertner, W., 1994. Distributive justice: theoretical foundations and empirical findings. *European Economic Review* 38, 711–772.
Greiner, B., 2004. An online recruitment system for economic experiments. In Kremer, K., Macho, V., *Forschung und wissenschaftliches Rechnen* GWDG Bericht 63, Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.
Grossman, Z., 2010. Strategic ignorance and the robustness of social preferences. Working paper, University of California at Santa Barbara.
Haisley, E.C., Weber, R.A., 2010. Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior* 68, 614–625.
Von Hippel, W., Trivers, R., 2011. The evolution and psychology of self-deception. *Behavioral and Brain Sciences* 34(1), 1–56.
Konow, J., 2001. Fair and square: the four sides of distributive justice. *Journal of Economic Behavior & Organizations* 46(2), 137-164.
Konow, J., 2000. Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review* 90(4), 1072-1091.
Konow, J., 2009. Is fairness in the eye of the beholder? An impartial spectator analysis of justice. *Social Choice and Welfare* 33(1), 101-127.
Konow, J., Saijo, T., Akai, K., 2009. Morals and Mores: Experimental Evidence on Equity and Equality. Mimeo.
Lazear, E.P., Malmendier, U., Weber, R.A., 2012. Sorting in Experiments with Application to Social Preferences. *American Economic Journal: Applied Economics* 4(1), 136-163.
Loewenstein, G., Issacharoff, S., Camerer, C., Babcock, L., 1993. Self-Serving Assessments of Fairness and Pretrial Bargaining. *Journal of Legal Studies* 22(1), 135-159.
Lonnqvist, J., Irlenbusch, I., Walkovitz, G., 2014. Moral hypocrisy: impression management of self-deception?, *Journal of Experimental Social Psychology*, 55, 53-62.
Monin, B., Merritt, A.. 2010. Moral hypocrisy, moral inconsistency, and the struggle for moral integrity. In M. Mikulincer and P. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil*, Herzliya Series on Personality and Social Psychology, 3, American Psychological Association.
Mui, V.-L.. 1995. The economics of envy. *Journal of Economic Behavior & Organizations* 26(3), 311-336.

Murnighan, J.K., Pillutla, M.M., 1995. Fairness versus self-interest:Asymmetric moral imperatives in ultimatum bargaining. In Kramer, R. M., and Messick, D. M. (Eds.), *Negotiation as a Social Process*, Sage, Thousand Oaks, CA.

Ostrom, E.. 2000. Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives* 14(3), 137-158.

Oxoby, R.J., Smith, A., 2012. Can Cognitive Dissonance Affect Social Preferences? Mimeo.

Rabin, M., 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83(5), 1281-1302.

Rabin, M.. 1994. Cognitive dissonance and social change. *Journal of Economic Behavior & Organization* 23(2), 177-194.

Rodriguez-Lara, I., Moreno-Garrido, L.J.B., 2012. Self-interest and fairness: self-serving choices of justice principles. *Experimental Economics* 15(1), 158-175.

Roth, A.E., Murnighan, J.K., 1982. The Role of Information in Bargaining: An Experimental Study. *Econometrica* 50(5), 1123-1142.

Schokkaert, E., Lagrou, L., 1983. An empirical approach to distributive justice. *Journal of Public Economics* 21, 33–52.

Stone, J., Wiegand, A.W., Cooper, J., Aronson, E., 1997. When exemplification fails: Hypocrisy and the motive for self-integrity. *Journal of Personality & Social Psychology* 72(1), 54-65.

Trivers, R., 2011. *Deceit and self-deception. Fooling Yourself the Better to Fool Others.* New-York, NY: Allen Lane, Penguin Books.

Valdesolo, P., DeSteno, D., 2008. The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology* 44 (5), 1334-1338.

Watson, G.W, Teague, B.T., Papamarcos, S.D.. 2006. Moral Hypocrisy: A Matter of Measure? In *Trends in Contemporary Ethical Issues,* ed. Aidan E. Wurtzel, 1-14. New-York, NY: Nova Science Publishers.

Zizzo, D.J., Oswald, A.J., 2001. Are People Willing to Pay to Reduce Others' Incomes? *Annales d'Economie et de Statistique* 63-64, 39-62.

Zizzo, D.J., 2010. Experimenter demand effects in economic experiments. *Experimental Economics* 13, 75-98.

Tables

Table 1: Mean fairness and unfairness judgments, for the three games in session 1

Judgments	In the shoes of Player A		In the shoes of Player B	
	Fairness	Unfairness	Fairness	Unfairness
Dictator game				
Lower bound	24.19 (1.79)	18.93 (1.68)	24.52 (1.53)	22.29 (1.98)
Upper bound	57.65 (1.53)	68.67 (1.76)	62.53 (1.70)	71.04 (1.81)
Ultimatum Bargaining game				
Lower bound	24.71 (1.56)	20.24 (1.58)	27.29 (1.58)	22.58 (1.57)
Upper bound	58.45 (1.36)	66.14 (1.86)	66.42 (1.84)	71.45 (2.0)
Trust game – Low initial transfer of 1 point				
Lower bound	20.33 (0.99)	17.47 (1.15)	16.11 (1.12)	13.40 (1.15)
Upper bound	51.36 (2.33)	56.78 (2.74)	43.67 (1.99)	52.26 (2.40)
Trust game – High initial transfer of 4 points				
Lower bound	34.30 (1.35)	29.55 (1.43)	29.34 (1.47)	25.80 (1.72)
Upper bound	65.34 (1.75)	67.97 (2.07)	56.98 (1.66)	63.86 (1.99)

Notes: The mean bounds in the Dictator and Ultimatum scenarios are expressed in percentage points of the initial endowment transferred by the sender to the receiver. For example, in the shoes of player A in the Dictator scenario people consider on average that fairness requires to transfer a minimum of 24.19% of the endowment to the receiver. The mean bounds in the Trust scenarios are expressed in terms of the percentage points of his total income (that is equal to 3 times the amount received from the sender + 5 points) returned by the receiver to the sender. These values include the data from all the participants in session 1, regardless of their actual role in session 2 (N = 83). Standard errors are in parentheses.

Table 2: Evolution of the mean fairness and unfairness judgments between the first and second sessions by actual role, for the three games.

A. Mean fairness judgments

Judgments	Actual Player A in session 2			Actual Player B in session 2		
	Session 1	Session 2	Diff. in %	Session 1	Session 2	Diff. in %
Dictator game						
Lower bound	22 (2.38)	15.80 (2.21)	-28.18	26.33 (2.64)	20.74 (2.14)	-21.23
Upper bound	59.44 (2.57)	56.34 (2.80)	-5.22	55.90 (1.67)	55.31 (2.33)	-1.06
Ultimatum Bargaining game						
Lower bound	25.24 (2.16)	24.31 (2.11)	-3.68	24.17 (2.27)	25.90 (2.15)	+7.16
Upper bound	58.83 (2.06)	58.83 (2.29)	0	58.05 (1.80)	61.37 (2.28)	+5.72
Trust game – Low initial transfer of 1 point						
Lower bound	17.68 (1.63)	15.55 (1.73)	-12.05	14.58 (1.53)	14.29 (1.51)	- 1.99
Upper bound	46.34 (3.03)	47.87 (3.65)	+ 3.30	41.07 (2.57)	44.35 (2.90)	+7.99
Trust game – High initial transfer of 4 points						
Lower bound	28.41 (2.46)	24.96 (2.00)	-12.14	30.25 (1.66)	26.75 (1.77)	-11.57
Upper bound	55.38 (2.58)	53.66 (2.49)	- 3.11	58.54 (2.11)	53.50 (1.98)	-8.61

B. Mean unfairness judgments

Judgments	Actual Player A in session 2			Actual Player B in session 2		
	Session 1	Session 2	Diff. in %	Session 1	Session 2	Diff. in %
Dictator game						
Lower bound	16.59 (2.14)	13.73 (1.98)	-17.24	21.21 (2.55)	16.95 (1.84)	-20.08
Upper bound	71.63 (2.58)	65.49 (2.93)	-8.60	65.79 (2.36)	62.57 (2.73)	-4.89
Ultimatum Bargaining game						
Lower bound	19.57 (2.14)	18.14 (2.06)	-7.31	20.92 (2.35)	17.24 (1.89)	-17.59
Upper bound	66.52 (2.83)	63.10 (2.47)	-5.14	65.76 (2.45)	64.15 (2.98)	-2.45
Trust game – Low initial transfer of 1 point						
Lower bound	11.28 (1.50)	13.11 (1.38)	+16.22	15.48 (1.69)	12.20 (1.78)	-21.19
Upper bound	52.13 (3.70)	54.27 (3.99)	+4.11	52.38 (3.11)	50.30 (3.23)	- 3.97
Trust game – High initial transfer of 4 points						
Lower bound	23.10 (2.65)	23.53 (2.01)	+1.86	28.43 (2.15)	25.21 (1.79)	-11.33
Upper bound	62.55 (3.07)	62.41 (2.36)	- 0.22	65.13 (2.56)	62.89 (1.92)	-3.44

Notes: In the Dictator and Ultimatum scenarios, the tables display the mean bounds reported by individuals in the position of player A (senders). These bounds are expressed in percentage points of the initial endowment transferred by the sender to the receiver. In the Trust scenarios, the tables display the mean bounds reported by individuals in the position of player B (receivers). These bounds are expressed in percentage points of his total income (that is equal to 3 times the amount received from the sender + 5 points) returned by the receiver to the sender. 'Diff. in %' means difference in percentages. Standard errors are in parentheses. (N = 42).

Table 3: Distribution of transfers by senders, hypothetical and real, in Dictator game and Ultimatum game

Transfer in points	Dictator game				Ultimatum game			
	Hypothetical choices		Actual choices		Hypothetical choices		Actual choices	
	#	%	#	%	#	%	#	%
0	6	14.63	14	34.15	0	0	0	0
1	2	4.88	5	12.20	3	7.14	2	4.76
2	9	21.95	11	26.83	5	11.90	5	11.90
3	11	26.83	5	12.20	14	33.33	14	33.33
4	9	21.95	2	4.88	12	28.57	12	28.57
5	4	9.76	3	7.32	7	16.67	9	21.43
6	0	0	0	0	1	2.38	0	0
7	0	0	0	0	0	0	0	0
8	0	0	1	2.44	0	0	0	0
Total	41	100	41	100	42	100	42	100

Note: In the Dictator game, there are 41 observations instead of 42 because one of the dictators participated only in the second session; therefore, his decisions are not taken into account in the data analysis. # indicates the number of observations for each possible transfer value, based on an endowment of 10 points.

Table 4: Distribution of amounts returned by receivers, hypothetical and real, in the Trust game

Amount returned in points	Low transfer of player A (1/5 point)				High transfer of player A (4/5 points)			
	Hypothetical choices		Actual choices		Hypothetical choices		Actual choices	
	#	%	#	%	#	%	#	%
0	3	7.14	16	38.10	0	0	10	23.81
1	12	28.57	17	40.48	1	2.38	2	4.76
2	22	52.38	8	19.05	2	4.76	3	7.14
3	4	9.52	0	0	1	2.38	5	11.90
4	0	0	1	2.38	4	9.52	7	16.67
5	1	2.38	0	0	5	11.90	6	14.29
6	-	-	-	-	7	16.67	3	7.14
7	-	-	-	-	6	14.29	6	14.29
8	-	-	-	-	11	26.19	0	0
9	-	-	-	-	4	9.52	0	0
10	-	-	-	-	0	0	0	0
11	-	-	-	-	1	2.38	0	0
Total	42	100	42	100	42	100	42	100

Note: # indicates the number of observations for each amount returned.

Table 5: Determinants of the adjustment of judgments, by scenario

Dependent variable: Variation of the lower bound of fairness between session 2 and session 1	Dictator	Ultimatum	Trust with low transfer	Trust with high transfer
Expectation on others' choice	.287 ***(.099)	.105 (.127)	.060 (.063)	.048 (.077)
Active role	1.403 (3.637)	-3.529 (2.362)	4.071 (2.503)	3.517 (3.333)
Difference (actual - hypothetical choice)	.239*** (.092)	.418*** (.117)	.240** (.108)	.274** (.131)
Constant	-12.508*** (3.285)	-1.564 (4.655)	-3.607 (2.252)	-4.624 (3.129)
N	83	83	83	83
R^2	.140	.132	.064	.054

Notes: All variables are expressed in percentage points of the endowment. Models are Ordinary Least Square models with robust standard errors in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level

Table 6: Comparison between fairness judgments in sessions 1 and 2, hypothetical and actual transfers, by game and by category of actual transfers

Percentage of variation	Actual – hypothetical transfers (1)	Judgment in session 2 – judgment in session 1 (2)	Actual transfer – judgment in session 1 (3)	Actual transfer – judgment in session 2 (4)
Dictator game				
Below median (N=19)	-86.49*** (<.001)	-34.65** (.028)	-88.97*** (<.001)	-83.12*** (.001)
Median (N=11)	-40.55*** (.007)	-46.73** (.018)	-19.71 (.165)	50.72** (.034)
Above median (N=11)	31.43 (.183)	17.72 (.521)	162.85*** (.005)	123.28*** (.003)
Total (N=41)	-33.06*** (.002)	-28.18** (.012)	-19.10 (.171)	12.66 (.498)
Ultimatum game				
Below median (N=21)	-14.03* (.076)	-18.59** (.036)	24.14* (.061)	52.49*** (<.001)
Median (N=12)	6.67 (.250)	-15.41 (.600)	46.79** (.017)	49.08*** (.003)
Above median (N=9)	25.00** (.017)	15.77 (.280)	50.01*** (.009)	30.44*** (.009)
Total (N=42)	2.07 (.683)	-3.68 (.284)	38.67*** (<.001)	43.97*** (<.001)
Trust game – Low initial transfer of 1 point				
Below median (N=16)	-100.00*** (<.001)	-6.00 (.371)	-100*** (<.001)	-100*** (<.001)
Median (N=17)	-41.52*** (.002)	-10.71 (.320)	-10.71 (.480)	0.00 (.367)
Above median (N=9)	-13.28 (.257)	14.10 (.390)	42.31 (.143)	24.72* (.085)
Total (N=42)	-49.42*** (<.001)	-2.56 (.537)	-24.79** (.044)	-22.81* (.094)
Trust game – High initial transfer of 4 points				
Below median (N=15)	-89.25*** (<.001)	-25.40* (.072)	-89.40*** (<.001)	-85.79*** (<.001)
Median (N=5)	-39.39** (.041)	-26.09 (.157)	-13.04 (.876)	17.65 (.670)
Above median (N=22)	-13.11*** (.002)	0 (.766)	18.66** (.028)	18.66** (.012)
Total (N=42)	-37.30*** (<.001)	-11.48* (.068)	-22.18* (.062)	-12.09 (.343)

Notes: The categories below median, median and above median refer to the actual transfers made in session 2. Table 6 only considers the lower bounds of fairness judgments. It displays the data relative to the players in the role of sender in the Dictator and Ultimatum games and to the players in the role of receiver in the Trust game. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level in Wilcoxon signed-rank tests.

Appendix A. Instructions for the first session (Translated from French)

We thank you for participating in this experiment that consists of two sessions. We remind you that you have committed to participate in the two sessions.

During these two sessions, you will be able to earn money. The amount of your earnings depends on your decisions and on the decisions of other participants in this experiment. Your earnings during these two sessions will be added up and paid to you at the end of the second session. This means that the earnings you will make during today's session will be paid to you at the end of the second session.

Throughout the two sessions, we will use points, with the following conversion rate between points and Euros:

1 point = 2 Euros

You will be paid in cash and in private in a separate room by somebody who is not aware of the content of this experiment. No other participant will be informed on your individual payoffs. Your answers will be always kept anonymous and confidential. You will never have to enter your name in the computer.

You have been given a tag indicating the name of a computer and a password. This password is strictly confidential and personal. Do not forget to keep this tag with you and to bring it back with you to be allowed to participate in the second session. If you lose your tag, you will not be allowed to participate in the second session and thus, you will not be paid at all.

Today, you will receive the instructions for the first session only and you will earn 8 Euros. This amount does not depend on your decisions. Please note that the content of the second session will not be affected by your decisions in today's session.

Today, the experiment consists of four independent parts. You have received a set of instructions for the first part. You will receive other sets of instructions once this part will have been completed.

Part 1

We ask you to answer to questions related to a scenario. This scenario is the following:

Imagine that a participant A receives 10 points that he can share with a participant B. A keeps for himself the points he has not transferred to B. B has no decision to make.

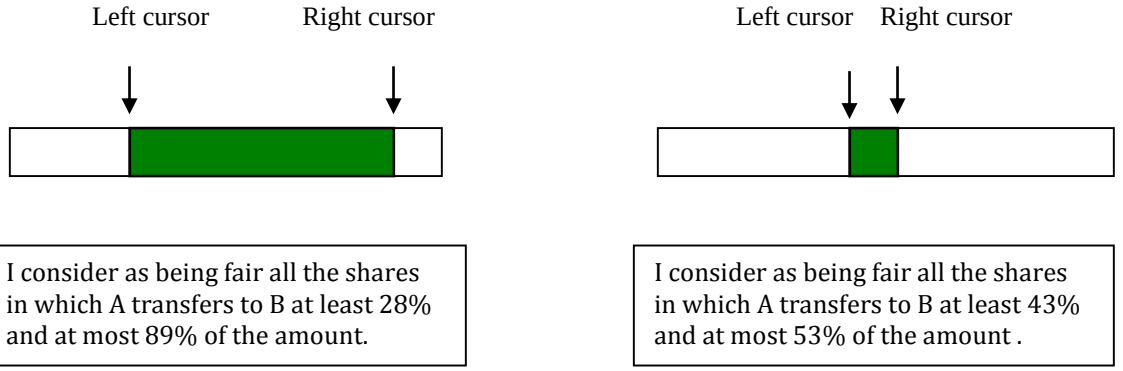
In this scenario, A earns: 10 points – the amount transferred to B.

B earns: the amount transferred by A.

We ask you to imagine first that you are the participant A and we ask you to answer the following questions:

1) What do you consider as being fair shares between A and B?

A horizontal bar will appear on your screen, together with two cursors, as indicated in the two following examples chosen at random.



* You move the left cursor to indicate the share of the amount transferred from A to B from which you consider the shares as being fair.

* You move the right cursor to indicate the share of the amount transferred from A to B up to which you consider the shares as being fair.

The dark inside area so defined indicates all the shares that you consider as being fair.

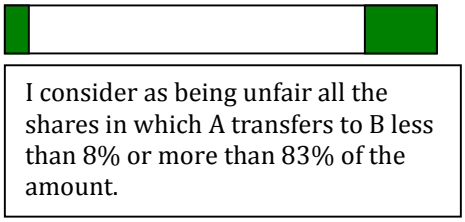
In the first example, all the shares in which A transfers to B at least 28% and at most 89% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 72% and a minimum of 11% of the received amount are considered as being fair.

In the second example, all the shares in which A transfers to B at least 43% and at most 53% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 57% and a minimum of 47% of the received amount are considered as being fair.

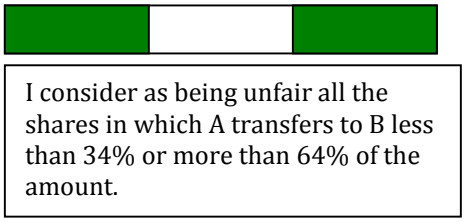
Next, you validate your answer by pressing the OK button. Once you have pressed this button, the following question will appear on your screen.

2) What do you consider as being unfair shares between A and B?

Here too, a horizontal bar with two cursors will appear on your screen, as indicated in the two following examples chosen at random.



I consider as being unfair all the shares in which A transfers to B less than 8% or more than 83% of the amount.



I consider as being unfair all the shares in which A transfers to B less than 34% or more than 64% of the amount.

* You move the left cursor to indicate the share of the amount transferred from A to B below which you consider the shares as being unfair.

* You move the right cursor to indicate the share of the amount transferred from A to B beyond which you consider the shares as being unfair.

The two outside dark areas so defined indicate the set of shares that you consider as being unfair. Then, you validate your choices by pressing the OK button.

In the first example, all the shares in which A transfers to B less than 8% and those in which he transfers more than 83% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 92% and those in which he keeps for himself less than 17% of the received amount are considered as being unfair.

In the second example, all the shares in which A transfers to B less than 34% and those in which he transfers more than 64% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 66% and those in which he keeps for himself less than 36% of the received amount are considered as being unfair.

Next, the following question will appear on your screen.

3) What do you think most people consider as being fair shares between A and B?

4) What do you think most people consider as being unfair shares between A and B?

Then, we will ask you the following questions:

5) If you could decide on the share between you and another participant, which amount would you decide to transfer to him?

6) If the other participants in today's session could decide on the share, which amount do you think would be transferred by others on average?

After you have responded to these questions, we will ask you to imagine that you are the participant B. Then, you will answer to the same questions regarding the definition of the fair and unfair shares from your own point of view, and then from the point of view of most people, according to you.

If you have any question regarding these instructions, please raise your hand. We will answer to your questions in private. During the session, communication between participants is strictly forbidden.

Before we start, please answer to the understanding questionnaire. We will check your answers individually. Then, you will enter your password and part 1 will begin.

Part 2

(*These instructions were distributed after the completion of part 1*)

We present you with a second scenario.

Imagine that a participant A receives 10 points that he can share with a participant B. A keeps for himself the points he has not transferred to B.

B can accept or reject A's offer. If B rejects A's offer, both A and B earn 0 (all the points are cancelled out). If B accepts A's offer, B earns the amount transferred by A and A keeps for himself the difference between the 10 points he has received and the amount transferred to B.

In this scenario, if his offer is accepted, A earns: 10 points – the amount transferred to B.

if his offer is rejected, A earns: 0 point.

if he has accepted A's offer, B earns: the amount transferred by A

if he has rejected A's offer, B earns: 0 point.

We ask you to imagine first that you are the participant A and we ask you to answer the following questions:

1) What do you consider as being fair shares between A and B?

2) What do you consider as being unfair shares between A and B?

3) What do you think most people consider as being fair shares between A and B?

4) What do you think most people consider as being unfair shares between A and B?

5) If you could decide on the share between you and another participant, which amount would you decide to transfer to him?

6) If the other participants in today's session could decide on the share, which amount do you think would be transferred by others on average?

Then, we will ask you to imagine that you are the participant B. Then, you will answer to the same questions regarding the definition of the fair and unfair shares from your own point of view, and then from the point of view of most people, according to you.

If you have any question regarding these instructions, please raise your hand. We will answer to your questions in private. Before we start, please answer to the understanding questionnaire. We will check your answers individually.

(*The following instructions were distributed after the completion of part 2*)

The following instructions are for part 3 and part 4.

Part 3

We present you with a new scenario that takes two versions (scenario 3 and scenario 3 bis).

Imagine that two participants, A and B, receive 5 points each. A can send to participant B a certain amount, in between 0 and 5 points. A keeps for himself the points he has not transferred to B. Each point sent to B is tripled. Then, B can send points back to A and he keeps for himself the points he has not sent back. In this scenario, A earns: 5 points – the

amount transferred to B + the amount sent back by B. B earns: 5 points + 3 times the amount transferred by A - the amount sent back to A.

Scenario 3. Imagine that A transfers 1 point to B and keeps 4 points for himself. Thus, B receives 3 points in addition to his initial 5 points. B can send back to A between 0 and 8 points.

We ask you to imagine first that you are the participant B and we ask you to answer the same questions as in the two previous scenarios regarding the definition of the fair and unfair amounts sent back by B to A, from your own point of view, and then from the point of view of most people, according to you. Then, we will ask you to imagine that you are the participant A. You will again answer to these questions about the definition of the fair and unfair amounts sent back to A.

Scenario 3 bis. Next, imagine that A transfers 4 points to B and keeps 1 point for himself. This, B receives 12 points in addition to his initial 5 points. B can send back to A between 0 and 17 points.

We ask you to answer to the same questions as for the scenario 3, first in imagining that you are the participant B and then as the participant A.

Part 4

This part will automatically start once part 3 has been completed. It consists of a questionnaire including 60 statements. For each statement, you must choose among five possible options the option that corresponds the most to your opinion.

Choose **SD** (Strongly Disagree) if the statement is absolutely wrong or if you strongly disagree.

Choose **D** (Disagree) if the statement is rather wrong or if you disagree.

Choose **N** (Neutral) if the statement is equally wrong or true or if you cannot choose or if you have no opinion.

Choose **A** (Agree) if the statement is rather true or if you agree.

Choose **SA** (Strongly agree) if the statement is absolutely true or if you strongly agree.

There is no "good" or "bad" answers. The aim of this questionnaire will be reached if you describe yourself and if you express your opinion as exactly as possible.

End of the session

Once you have completed part 4, a few last questions will appear on your computer screen; then you will receive a message allowing you to leave the lab. Between the beginning of this questionnaire in part 4 and the moment you will be allowed to leave the lab, 10 minutes minimum will have elapsed.

Last, please give us back the instructions and do not forget to take your tag with your password to be allowed to participate in the second session.

We thank you for not communicating with anyone about the questions and your answers in this session. **It is indeed very important that nobody else is aware of your answers in this session.**

Please read again these instructions. If you have any question, please raise your hand. Before we start part 3, please answer to the understanding questionnaire.

Appendix B. Instructions for the second session

We thank you for participating in this second session that consists of four parts. Today, you will be able to earn additional payoffs. The amount of your earnings depends on your decisions and on the decisions of other participants. Your earnings will be calculated in points, with the following conversion rate between points and Euros:

1 point = 2 Euros

The payoffs earned in each of the two sessions will be added up and converted into Euros at the end of this session. The total amount of your earnings will remain confidential. We remind you that you will be paid in cash and in private in a separate room by somebody who is not aware of the content of this experiment.

The identity of the participants with whom you will interact during this session will never be communicated to you. Your answers will be kept anonymous and confidential. For this reason, we ask you not to communicate your choices to anybody during or after the experiment.

One of the three first parts of this session will be randomly drawn at the end of the session and you will be paid what you have actually earned in this part. The random draw is independent for each participant. Moreover, you will receive €8 for your participation in today's session that will be added up to the €8 that you have already earned last week.

Before we start the first part, we ask you to enter your password in your computer and to answer to a preliminary question.

Part 1

You are randomly matched with another participant. One of you is the "participant A", the other is the "participant B". The assignment of roles is random.

If you are a participant B, you have no other decision to make during this part. What occurs depends only on participant A.

If you are a participant A, you receive 10 points. You must decide on the amount, between 0 and 10 points, that you are willing to transfer to the participant B and you keep the rest for yourself. Once you have made your choice, you press the OK button to validate your choice.

You make your decision once.

The payoffs are determined as follows:

The participant A earns: 10 points – the amount transferred to B
The participant B earns: the amount transferred from A

A feedback on the amount transferred by A to B will be given to B only at the end of the session. At the end of the session, you will also be informed on whether this part has been selected for payment in Euros.

Once A has made his decision, both A and B can see a question displayed on their computer screen. A correct answer to this question will allow you to earn €1 more.

We remind you that communication between participants is strictly forbidden. If after reading these instructions again you have any question, please raise your hand. We will answer to your questions in private. Please fill out the understanding questionnaire that has been distributed.

Part 2 (*The following instructions were distributed after the completion of part 1*)

In part 2, you are randomly matched with another participant. This co-participant is likely another person than in the previous part. One of you will be "participant A", the other one will be the "participant B". The assignment of roles is random and independent of the previous part.

If you are a participant A, you receive 10 points. You make an offer to the participant B about the amount, between 0 and 10 points, that you are willing to transfer to him. Once you have made your decision, please press the OK button.

You make your decision once.

If you are a participant B, you decide on whether you accept or you reject the offer made by the participant A. However, you will not be informed immediately on the offer made by A. Your computer screen will display all the possible choices made by A and you will have to decide for each possible choice made by A if you accept or you reject it. Once you have made your series of decisions, please press the OK button.

What are the consequences of accepting or rejecting an offer?

If A's offer is rejected, both A and B earn 0 point.

If A's offer is accepted, B earns the amount transferred by A and A keeps for himself the difference between the 10 points he received initially and the amount transferred to B.

At the end of the session, we will match the amount actually offered by A to B and B's decision for this amount. Payoffs are calculated as follows.

Participant A:
If his offer is accepted, A earns: 10 points – the amount transferred to B
If his offer is rejected, A earns: 0 point.
Participant B:
If he has accepted A's offer, B earns: the amount transferred by A
If he has rejected A's offer, B earns: 0 point.

At the end of the session, we will inform B about the offer actually made by A and we will inform A about the decision of B for this offer. At the end of the session, you will also be informed on whether this part has been selected for payment in Euros.

Once A and B have made their decisions, both A and B can see a question displayed on their computer screen. A correct answer to this question will allow you to earn €1 more.

If after reading these instructions again you have any question, please raise your hand. We will answer to your questions in private. Please fill out the understanding questionnaire that has been distributed.

Part 3 (*The following instructions were distributed after the completion of part 2*)

In part 3, you are randomly matched with another participant. This co-participant is likely another person than in the previous parts. One of you will be "participant A", the other one will be the "participant B". The assignment of roles is random and independent of the previous parts.

Both participants A and B receive an initial endowment of 5 points.

If you are a participant A, you send to the participant B an amount, comprised in between 0 and 5 points, included, that is taken out of your endowment. Once you have made your decision, please press the OK button.

You make your decision once.

Once you have validated your decision, each point sent to B is tripled.

If you are a participant B, you decide on how many points you want to send back to the participant A, between 0 and your total number of points available (i.e. 5points + the tripled amount of points sent by A).

However, you will not be informed immediately on the amount actually sent by A. Your computer screen will display all the possible choices made by A. Then, you will have to decide for each possible amount sent by A how many points you want to send him.

This means that as a participant B, you must make several decisions regarding the amount you are willing to send back, one for each possible amount sent by A. Once you have made your series of decisions, please press the OK button.

At the end of the session, we will match the amount actually sent by A to B and the corresponding amount sent back by

B. Payoffs are calculated as follows.

The participant A earns: 5 points – the amount sent to B + the amount sent back by B
The participant B earns: 5 points + 3 times the amount sent by A - the amount sent back to A

At the end of the session, we will inform B about the amount actually sent by A and we will inform A about the corresponding amount actually sent back by B. At the end of the session, you will also be informed on whether this part has been selected for payment in Euros.

Once A and B have made their decisions, both A and B can see a question displayed on their computer screen. A correct answer to this question will allow you to earn €1 more. We consider an answer as being correct if is exact at 10%.

If after reading these instructions again you have any question, please raise your hand. We will answer to your questions in private. Please fill out the understanding questionnaire that has been distributed.

Part 4 (*The following instructions were distributed after the completion of part 3*)

We will present three scenarios on your computer screen successively (one of which has two versions) in which a participant A and a participant B interact together. Each scenario replicates exactly the rules of each of the three previous parts. The scenario 1 corresponds to part 1, the scenario 2 corresponds to part 2 and the scenarios 3 and 3 bis correspond to part 3.

In each scenario, you are requested to imagine that you are the participant A and we ask you the following questions.

1) What do you consider as being fair shares between A and B or fair amounts sent back from B to A (according to the scenario)?

2) What do you consider as being unfair shares between A and B or unfair amounts sent back from B to A (according to the scenario)?

3) What do you think most people consider as being fair shares or fair amounts sent back?

4) What do you think most people consider as being unfair shares or unfair amounts sent back?

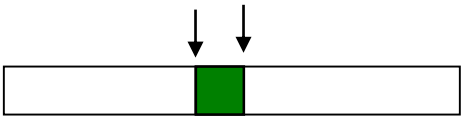
Then, we will ask you to imagine that you are the participant B and you will answer to the same questions. In addition, participants B in parts 1 and 2 and participants A in part 3 will have to answer to an additional question in each scenario.

To enter your answers about the fair shares or amounts sent back, a horizontal bar with two cursors will appear on your screen, as indicated in the two following random examples.

Left cursor Right cursor

I consider as being fair all the shares in which A transfers to B at least 28% and at most 89% of the amount.

Left cursor Right cursor



I consider as being fair all the shares in which A transfers to B at least 43% and at most 53% of the amount .

* You move the left cursor to indicate the share of the amount transferred from A to B or the amount sent back from B to A from which you consider the share or the amount sent back as being fair.

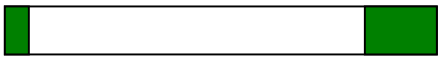
* You move the right cursor to indicate the share of the amount transferred from A to B or the amount sent back from B to A up to which you consider the share or the amount sent back as being fair.

The dark area so defined indicates all the shares or the amount sent back that you consider as being fair.

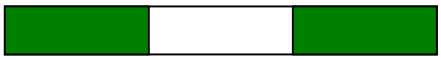
In the first example, all the shares in which A transfers to B at least 28% and at most 89% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 72% and a minimum of 11% of the received amount are considered as being fair.

In the second example, all the shares in which A transfers to B at least 43% and at most 53% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 57% and a minimum of 47% of the received amount are considered as being fair.

To enter your answers about the unfair shares or amounts sent back, you also use the horizontal bar with the two cursors, as indicated in the two following random examples.



I consider as being unfair all the shares in which A transfers to B less than 8% or more than 83% of the amount.



I consider as being unfair all the shares in which A transfers to B less than 34% or more than 64% of the amount.

* You move the left cursor to indicate the share of the amounts transferred from A to B or the amounts sent back by B to A below which you consider the shares or the amounts sent back as being unfair.

* You move the right cursor to indicate the share of the amounts transferred from A to B or the amounts sent back by B to A beyond which you consider the shares or the amounts sent back as being unfair.

The two outside dark areas so defined indicate the set of shares or amounts sent back that you consider as being unfair.

In the first example, all the shares in which A transfers to B less than 8% and those in which he transfers more than 83% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 92% and those in which he keeps for himself less than 17% of the received amount are considered as being unfair.

In the second example, all the shares in which A transfers to B less than 34% and those in which he transfers more than 64% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 66% and those in which he keeps for himself less than 36% of the received amount are considered as being unfair.

End of the experiment

At the end of the fourth part, we will give you a feedback on the actual choice of your co-participant in each of the first three parts and on your associated potential payoffs. Then, we will randomly draw the part that will be used for your actual payment.

After answering last questions, you will be invited to leave the room.

Appendix C. Tables

Table A1: Determinants of hypothetical choices, by scenario

Dependent variable: Hypothetical choice in session 1	Dictator Tobit	Ultimatum OLS	Trust with low transfer Tobit	Trust with high transfer Tobit
Lower bound of fairness	.293** (.118)	.401*** (.117)	.371** (.150)	-.155 (.191)
Upper bound of fairness	.050 (.113)	.263*** (.099)	.116 (.081)	.055 (.126)
Expectations about others'				
- hypothetical choice	.761*** (.128)	.327*** (.114)	.538*** (.118)	.404** (.179)
- lower bound of fairness	-.325*** (.119)	.003 (.118)	-.012 (.150)	.225 (.173)
- upper bound of fairness	-.092 (.108)	-.032 (.080)	-.074 (.091)	.055 (.126)
Constant	9.020 (7.432)	-1.313 (4.973)	1.192 (3.807)	19.906*** (4.966)
N	83	83	83	83
Left-censored obs.	11	-	10	2
Log pseudolikelihood	-290.787	-	-282.883	-307.300
R^2	.078	.544	.067	.051

Notes: All variables are expressed in percentage points of the endowment. We use OLS instead of Tobit in the Ultimatum game since no data were censored. Robust standard errors in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A2: Determinants of actual choices, by scenario

Dependent variable: Actual choice in session 2	Dictator Tobit	Ultimatum OLS	Trust with low transfer Tobit	Trust with high transfer Tobit
Hypothetical choice	.827*** (.256)	.241** (.102)	.404*** (.130)	.610*** (.223)
Expectations about others' actual choice	-.034 (.235)	.663*** (.147)	.719*** (.164)	.682*** (.177)
Constant	-9.758 (7.432)	2.286 (4.500)	-23.545*** (6.553)	-22.571*** (8.253)
N	41	42	42	42
Left-censored obs.	14	0	16	10
Log pseudolikelihood	-134.027	-	-105.848	-137.709
R^2	.030	.634	.142	.106

Notes: All variables are expressed in percentage points of the endowment. We use OLS instead of Tobit in the Ultimatum game since no data were censored. Robust standard errors in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level